

UniControl: A Unified Diffusion Model for Controllable Visual Generation In the Wild

Year: 2023 | Citations: 187 | Authors: Can Qin, Shu Zhang, Ning Yu

Abstract

Achieving machine autonomy and human control often represent divergent objectives in the design of interactive AI systems. Visual generative foundation models such as Stable Diffusion show promise in navigating these goals, especially when prompted with arbitrary languages. However, they often fall short in generating images with spatial, structural, or geometric controls. The integration of such controls, which can accommodate various visual conditions in a single unified model, remains an unaddressed challenge. In response, we introduce UniControl, a new generative foundation model that consolidates a wide array of controllable condition-to-image (C2I) tasks within a singular framework, while still allowing for arbitrary language prompts. UniControl enables pixel-level-precise image generation, where visual conditions primarily influence the generated structures and language prompts guide the style and context. To equip UniControl with the capacity to handle diverse visual conditions, we augment pretrained text-to-image diffusion models and introduce a task-aware HyperNet to modulate the diffusion models, enabling the adaptation to different C2I tasks simultaneously. Trained on nine unique C2I tasks, UniControl demonstrates impressive zero-shot generation abilities with unseen visual conditions. Experimental results show that UniControl often surpasses the performance of single-task-controlled methods of comparable model sizes. This control versatility positions UniControl as a significant advancement in the realm of controllable visual generation.